Technical Document 2455
February 1993

Machine Visual Targeting Modeled on Biological Reflexes

M. R. Blackburn

Technical Document 2455
February 1993

Machine Visual Targeting Modeled on Biological Reflexes

M. R. Blackburn

NAVAL COMMAND, CONTROL AND OCEAN SURVEILLANCE CENTER
RDT&E DIVISION
San Diego, California 92152-5000

J. D. FONTANA, CAPT, USN
Commanding Officer

R. T. SHEARER
Executive Director

ADMINISTRATIVE INFORMATION

The work in this document was performed in the Undersea Artificial Intelligence and Robotics Branch (Code 943) as part of the Independent Research Program of the Naval Command, Control and Ocean Surveillance Center RDT&E Division. Sponsorship was provided by the Cognitive and Neural Sciences Division of the Office of Chief of Naval Research, Arlington, VA.

Released by
P. J. Heckman, Jr., Head
Undersea AI and Robotics Branch

Under authority of
N. B. Estabrook
Ocean Engineering
Division

ACKNOWLEDGMENTS

The author gratefully appreciates the reviews by and discussions with Hoa Nguyen and Pat Shoemaker of the NCCOSC RDT&E Division.

PK

BACKGROUND

Visual target detection is a first step in target identification and the visual control of behavior in both natural and artificial visual/motor systems. This paper presents a simple algorithm for machine visual target detection and localization based on aspects of the known functional neuroanatomy of biological systems. The system can generate, from the relative motion of a target image on the sensor surface, a series of motor commands that reorient the high-resolution portion of the sensor surface to regions of the target that potentially contain the most information. The reorientations are analogous to saccadic eye movements of biological visual systems.

After light detection, one of the most primitive and phylogenetically persistent of visual functions is motion analysis. Motion analysis is used by insects (Wehner, 1981) and their amphibian predators (Ewert, 1978) for target detection and navigation. Similar mechanisms are most likely used for similar purposes in advanced mammals.

The Y ganglion cells of the vertebrate retina respond transiently to rapid change in contrast (Cleland et al., 1971), and, therefore, are considered to be suitable motion detectors (Breitmeyer, 1983). This activity is sent to the superior colliculus (SC) of the brain-stem tegmentum (or to its comparative region, the optic tectum in other vertebrates [Fite & Scalia, 1976]). The superior colliculus is organized retinotopically (Cynader & Berman, 1972). Straschill and Rieger (1973) were able to evoke eye movements through electrical stimulation of the superior colliculus (SC). The targets of the evoked eye movements were regions of the visual space that normally were represented by afferents to the area of the superior colliculus receiving the electrical stimulation. After integration and comparisons of input, the SC produces a specific set of signals to motor nuclei. These signals reflexively initiate commands to the eye, neck, and trunk muscles (via additional brain stem and spinal nuclei) to orient the sensor or the entire organism to the region of space where the dominant activity occurred.

We, therefore, created a neural network architecture based on anatomical and physiological evidence from the literature on the mammalian Y retina and superior colliculus. The algorithms determine competitively the region of greatest activity from the sensor surface and return coordinates to alter the fixation point to bring that region onto the high-resolution center of the sensor surface.

MODEL MECHANISMS OF TARGET DETECTION AND LOCALIZATION

The computational model of the Y retina used in this report is based on previous work (Blackburn et al., 1987), but is reduced in complexity to provide only for motion detection. The model is further modified to provide for graded variation in receptive field size with eccentricity (distance from the center of the receptor array) and to accomplish a polar to rectangular mapping from receptor to computational fields (Blackburn, 1993).

The receptor matrix is an N by N rectangular field (N = 256), indexed by x and y, compatible with the output of a video frame grabber. The receptive fields, composed of groups of neighboring receptors, are mapped to a smaller n by m rectangular matrix (n = m = 64), indexed by i and j, that becomes the ganglion output layer. The mapping is a log-polar transformation based on receptive field center locations and sizes that are computed once and stored in a lookup table. In mammalian retina, Y ganglion cells are absent near the visual field center, but appear with decreasing concentration from a few degrees eccentricity into the far periphery (Perry et al., 1984).

MOTION DETECTION IN THE RETINAL MODEL

In the current model of the Y retina, outlined in figure 1, the horizontal elements rapidly assume the potentials of their overlying receptors, and then counter receptor input to the underlying bipolar elements. These horizontal elements are modeled after the small *B* type described by Sterling (1983) that have no gap junctions, and synapse only with rods and rod bipolars. Both on-center and off-center bipolar elements are computed. The rapid matching of receptor potentials by the horizontals causes the bipolars to respond primarily to change in receptor potential and cease to respond to constant contrast spatially or temporally.

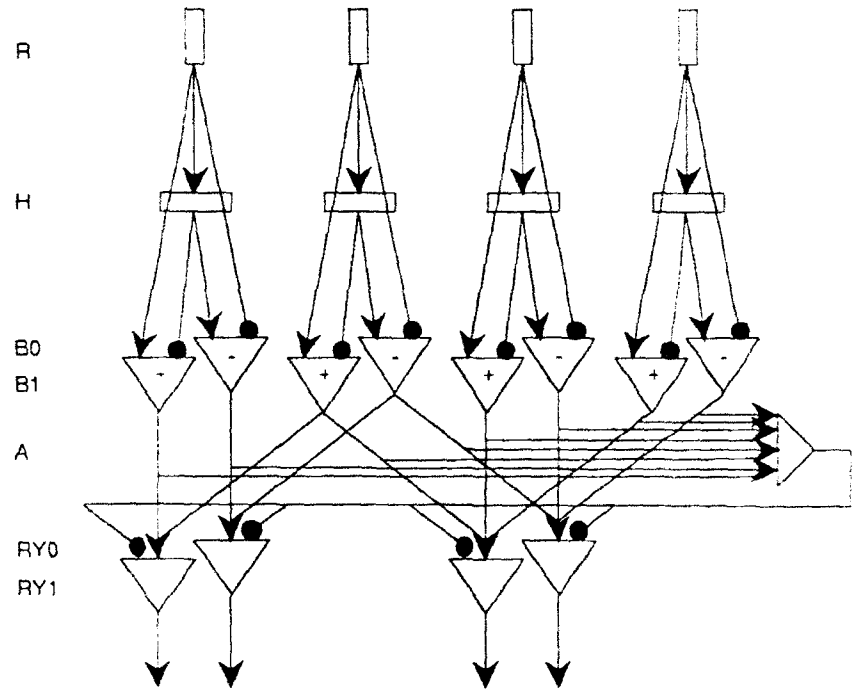


Figure 1. Model of Y retina for motion detection showing receptors (*R*), horizontals (*H*), on-center bipolars (*B1*), off-center bipolars (*B0*), amacrine (*A*), on-center ganglion (*RY1*), and off-center ganglion (*RY0*) elements. Only origins and terminals of lines make connections. Inhibition is indicate by a terminal dot. Excitation is indicated by a terminal arrow.

The amacrine element modulates bipolar to ganglion transmission to provide a consistent mean level of activity from the Y retina irrespective of light intensity and contrast. Individually, the Y ganglion elements are jiggle detectors, responding to temporal change in contrast within their receptive fields. The output of the motion analysis subnetwork of the retinal model is an analog potential from each ganglion element that represents, at any moment in time, the complexity of the image moving in the receptive field of that element.

The input potential on the horizontal (*Hb*) is the new receptor potential

$$Hb_{x,y} = R_{x,y} \ . \qquad [1]$$

where R is the receptor potential,and x and y locate the receptor on the sensor surface.

In one time step (t), the horizontal potential equilibrates to the input potential.

$$H_{x,y(t)} = Hb_{x,y(t-1)}. \qquad [2]$$

The bipolar elements then take the difference between the overlying photoreceptor and the nearest horizontal element. In one case, the bipolar elements pass the excess of the horizontal activity over the receptor average, while in the other case the bipolar elements pass the excess of the receptor activity over the horizontal. Bipolar elements do not pass negative potentials (they are half-wave rectified).

On-center bipolar activity ($B1$) is the greater of the receptor potentials over the equilibrated local horizontal,

$$B1_{x,y} = \max \{0, R_{x,y} - H_{x,y}\}. \qquad [3]$$

Off-center bipolar activity ($B0$) is the greater of the equilibrated local horizontal over the receptor potentials,

$$B0_{x,y} = \max \{0, H_{x,y} - R_{x,y}\}. \qquad [4]$$

The amacrine element (A) is driven toward the potential on the most active bipolar, then dissipates slowly.

$$A_{(t)} = K1 * \max \{A_{(t-1)}, B_{(t)}\}, \qquad [5]$$

where $K1 = 0.9$.

The on-center (*RY1*) and off-center (*RY0*) ganglion output elements are the sum of their respective bipolars in a local receptive field whose radius increases with eccentricity

$$RY0_{i,j} = (K2/A)^* \sum_{h,h} BO_{x\pm h,y\pm h}$$

$$RY1_{i,j} = (K2/A)^* \sum_{h,h} BI_{x\pm h,y\pm h}, \qquad [6]$$

where *K2* is half the maximum possible input potential (=128), and *h* ranges from 0 to the receptive field radius (Blackburn, 1993). This information is passed to the input layer of the model superior colliculus.

SACCADE GENERATION IN SUPERIOR COLLICULUS MODEL

The SC model has the surface dimensions corresponding to the coordinates of the Y retinal map (*m* by *n*). The vertical organization of the SC is given in figure 2.

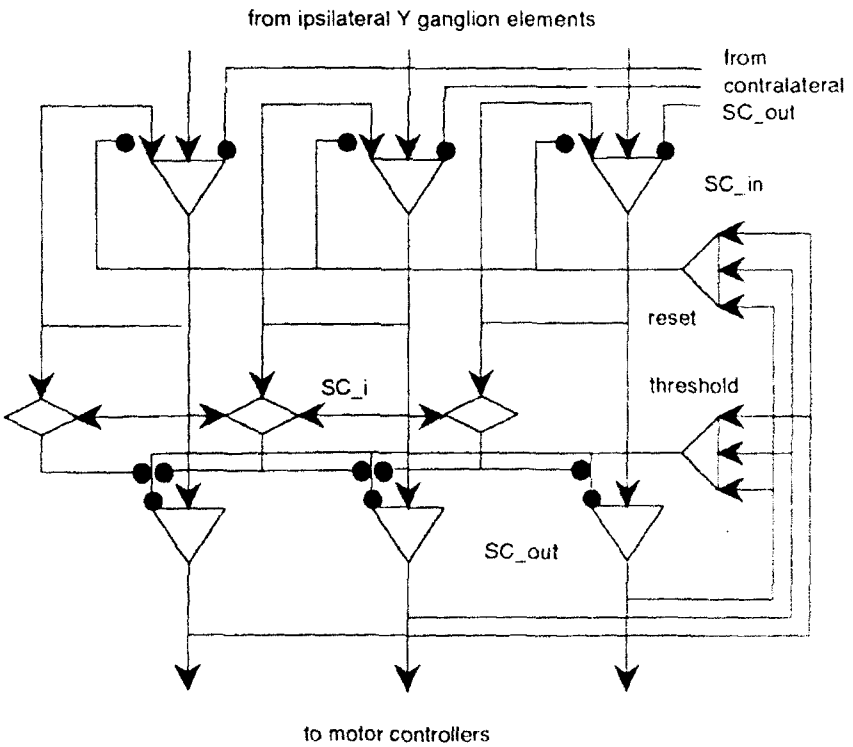


Figure 2. Vertical organization of SC. Input from the Y retina is temporally integrated by the *SC_in* elements. The SC_i elements perform a local spatial summation and the first *SC_i* element on any cycle that exceeds the threshold element potential opens the transmission from the *SC_in* element to the *SC_out* element. The potentials from the *SC_out* element then reset the threshold. Inhibition is indicated by dots, excitation is indicated by arrows. Crossing lines do not interconnect.

Because Rashbass (1961) found that saccades were not made to foveal targets within a visual angle of 0.3 degree from the center of the fovea, and because Y ganglion cells are generally absent in the fovea, we exclude the most central pixels of the central receptor surface (fovea) from the target competition.

A global threshold (Θ), maintained within the SC, is distributed to all output layer elements and generally prevents an output saccade command from being made. Eventually, however, the potentials on the intermediate layer elements accumulate and one reaches a value that exceeds the threshold. At that time, the inhibition from the threshold element to a select group of output elements associated with the winning intermediate layer element is blocked by the winning intermediate layer element.

The two bilateral superior colliculi are proposed to have a mutually inhibitory relationship. This inhibition is topographic on the SC surfaces. Thus, one output element in the SC inhibits one input element in the contralateral SC that represents a point on the retina exactly equidistant on a line running from the point represented by the first through the center of the retina. This relationship is illustrated in figure 3. The log-polar mapping of the Y retinal to the SC greatly simplifies the topographic inhibitory relationships between the bilateral SC. *The result of the mapping is that after a saccade, locations in the SC that were active during the last saccade command represent points on the retinal surface that are always contralateral to points that subsequently overlay the last saccade target.* This consequence of natural neural network geography allows an efficient means to prevent a return saccade to the last target. The two bilateral superior colliculi simply and mutually inhibit homologous points on the log-polar projection. Without such an inhibition, the saccades could become locked in a two-point limit cycle between the two most attractive features in the visual space. Without the log-polar projection, the routing of inhibitory fibers between the two SC would be considerably more complex.

The input to the SC from the Y retinal ganglion elements is retinotopically distributed and integrated over time, allowing excitation to build up in a local area if there is no or relatively little inhibition from the contralateral SC,

$$SC\_in_{m,n(t)} = K4*SC\_in_{m,n(t-1)} + RY0_{m,n(t)} + RY1_{m,n(t)} - SC\_out_{m,o(t-1)}, \quad [7]$$

where o indexes the homologous point in the contralateral SC $SC\_out_{m,o(t-1)}$. K_4 is a constant < 1.0 that determines the decay of the SC input $SC\_in_{m,n(t)}$.

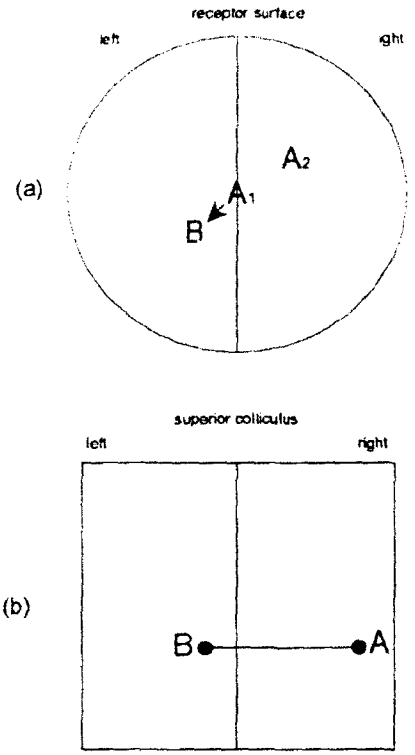


Figure 3. Mutual inhibition between homologous points of the superior colliculus prevents return saccades to the last target. When a saccade is made to target B in (a), the last target A1 becomes relocated at A2 on the receptor surface. Because locations A2 and B are equidistant from the center of the receptor surface, and lie upon the same diagonal, they also occupy homologous points on the bilateral log-polar map to the superior colliculus (b).

The input activity diffuses locally to the six nearest neighbors in the intermediate layer ($SC\_i_{m,n}$)

$$SC\_i_{m,n} = K_5 * \sum SC\_in_{m\pm z, n\pm z} \quad . \qquad [8]$$

where K_5 is a constant of diffusion < 1.0 and z is the radius of the local neighborhood.

The locally spatially and temporally summed SC input potentials ($SC\_i_{m,n}$) are compared to the global threshold. The order of the comparison is performed randomly between the two bilateral SC to remove any bias in sampling the intermediate layer as a result of the serial computation. The intermediate layer element with the largest potential that exceeds threshold is able to pass the SC_in activity out of the SC.

$$SC\_out_{I\pm z,J\pm z} = \begin{cases} SC\_in_{(I=m)\pm z,(J=n)\pm z},\ SC\_i_{m,n} > \Theta \\ 0, \text{ else.} \end{cases} \quad [9]$$

where I and J are the coordinates of the center of a small group of winning elements that lie within a radius z of the center, and Θ is the threshold.

The threshold is then reset to the sum of the winning output elements

$$\Theta = \Sigma SC\_out_{I\pm z,J\pm z}. \quad [10]$$

All other output elements remain at zero, and all input elements are reset to zero.

The threshold slowly decays on each program cycle,

$$\Theta_{(t)} = K_3 * \Theta_{(t-1)}, \quad [11]$$

where K_3 is a constant < 1.0.

The new target coordinates are determined by the location of the winning elements in the input field and their levels of activity. The final output is the vector sum of these contributions as has been described for the biological superior colliculus (Lee et al., 1988). This computation is presumably carried out in biological systems in retinal rectangular coordinates in the motor nuclei that receive SC output. In the present implementation, the changes in fixation in terms of retinal x and y coordinates required to foveate the new target is given by

$$\Delta x = (K_5/SC\_i_{I,J})*\Sigma\ (X_{I\pm z,J\pm z}\text{-}center)*SC\_out_{I\pm z,J\pm z}$$
$$\Delta y = (K_5/SC\_i_{I,J})*\Sigma\ (Y_{I\pm z,J\pm z}\text{-}center)*SC\_out_{I\pm z,J\pm z}. \quad [12]$$

where $X_{I\pm z,J\pm z}$ and $Y_{I\pm z,J\pm z}$ are the retinal map coordinates of the winning SC elements recovered from the lookup table, and K_5 is as defined for equation [9].

In the present implementation, Δx and Δy are sent to two virtual motors that move the point of fixation on the X and Y axes through the visual field. The SC_out potentials of the neighborhood that wins the competition to control the saccade are multiplied by the difference between their transformed coordinates and the center of the receptor field. These differences can be considered as fixed weights for the X and Y directions that must be traversed to target the region responsible for the winning activation. Those sums of

products are adjusted by the neighborhood sum (SC_I) to achieve the weighted vector average for movement on X or Y.

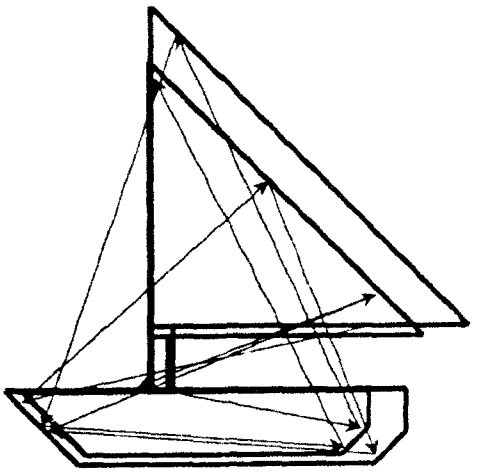
The present system has a very simple effector capability, limited to selecting a different portion of the video frame to pass to the retina. Therefore, movement is currently performed electronically within the limits of the video frame. The video camera provides a 512 by 512 pixel image encompassing about 22 degrees of visual angle. A 256 by 256 sample of the image is passed to the retina.

In the process described so far, the Y retina detects image motion of a moving target. The Y retina then sends this information to the SC where the most active region of the visual field is determined, and coordinates for motor commands to bring the center of the receptor surface upon that region of visual space are computed. This system will, however, detect potential targets only if the targets are moving. Static targets will be ignored. A simple solution to detect static targets has been developed for higher vertebrates that can easily move their eyes. Eye movement creates a relative motion of static objects on the retina. This allows the detection and localization of immobile animate objects as well as inanimate objects. The correlation between the ability to move the eyes and the presence of a high resolution fovea, reported by Ali and Klyne (1985), is not unexpected.

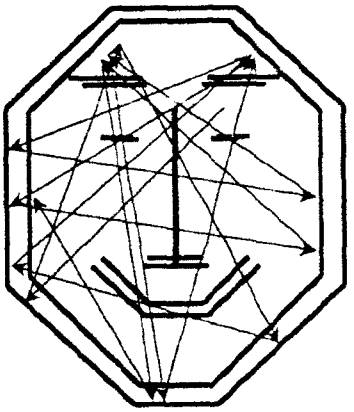
In the present model, when the threshold decreases to below 50% of its post saccade reset value, small oscillations in fixation occur. These oscillations are sufficient to cause the motion analysis subnetwork of the retina to detect the relative motion of regions of contrast on the receptor surface. Since this output is proportional to the distribution of contrast among the receptive fields of the Y ganglion elements, a competition for high-contrast regions in the peripheral visual field is possible. In the event that the immediate visual field contains no detectable contrast, the internal noise of the SC will eventually generate a saccade command to a random location as the threshold further drops.

BEHAVIOR OF THE MODEL

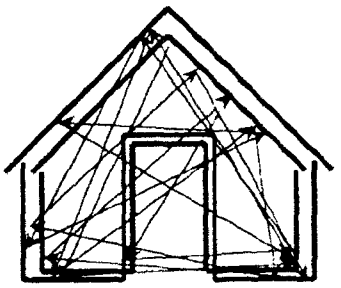
Scan paths to moving patterns are given in figure 4 a-c. The results reported here were obtained in numerical simulations on a Convex C-240 minisuper computer. Video camera input generates similar behavior. The sizes of the patterns were slowly increased and decreased about a common point, simulating motion toward or away from the observing receptor surface. The overlaid pairs of patterns in the figures reflect the size limits within which the patterns were varied. The change in size was about 15% on any dimension. The targets of each saccade are indicated by the arrowhead locations, and the order of saccades are indicated by the connecting lines. Most saccade targets were located on the perimeters of the patterns. This occurred because the targeting mechanism was sensitive only to contrast, and because the peripheral receptor surface was more sensitive to contrast than was the central receptor surface due to the larger peripheral receptive fields over which potentials were integrated. Longer saccades were thus favored over shorter saccades when a potential target was present. While some regions of the patterns attracted more saccades than others, target locations varied considerably.



(a)



(b)



(c)

Figure 4. Scan paths to moving patterns (a-c).

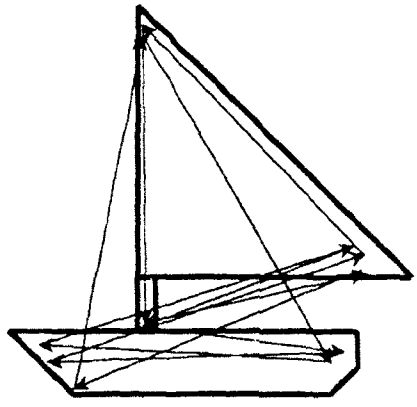
Scan paths for three static patterns are shown in figure 5 a-c. Regions of the image field were detected by the spontaneous jitter of the receptor surface that created a relative motion of the image. The figures show that within a sequence of less than 12 saccades, most of the high-information regions of the patterns were placed on the high-resolution portion of the receptor surface.

The scan paths for static patterns discovered by automotion (jitter) are very similar to the scan paths evoked by moving stimuli (figure 4). The automotion generated saccades did visit the centers of the patterns more often than the evoked saccades, but this is likely an artifact of the algorithm that changed the size of the images. In the dynamic condition, lines located at short distances from the expansion point changed location less often than lines located at long distances because of the floating point to integer conversions required in the algorithm, whereas all lines were moved in the jitter generation algorithm during each opportunity.

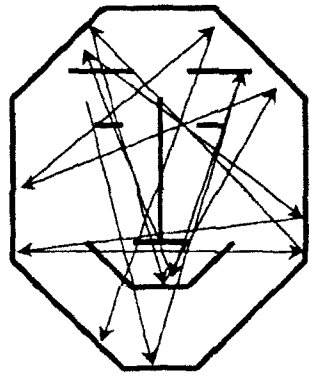
SCALE INVARIANCE

One possible way to achieve scaled output is through topographical mapping. When the motor map is isomorphic with the sensory map, an activation in the sensory plane will automatically be scaled to the motor response appropriate to the location of the stimulus. Thus, a large input pattern will produce a large response and a small input will evoke a small response. Figure 6 shows the scan path to an image that is twice the size of the image in figure 5b. While the sequence of targets differed between the scan paths of figure 5b and figure 6, many of the same targets are visited. Importantly, for the issue of scale invariance, the difference in image size did not change the relative importance of target regions.

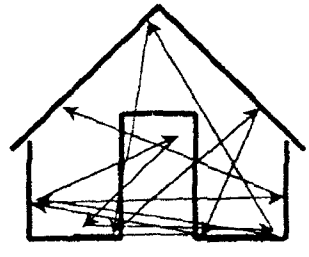
The SC in the present model is intended to perform functions of the mammalian SC, that is, to orient the subject to potentially significant visual stimuli. To do this, the SC must integrate relevant information and organize a response that would properly reposition the point of visual fixation. The requirement is to align the center of the receptor surface with the center of mass of the most attractive region of the visual field. The most attractive region is defined in the present algorithm as that region of the visual field that contains the greatest contrast. The detected level of contrast is a function both of the complexity of the local pattern of bright and dark regions and of its motion. The horizontal layer of the Y retina washes out any contrast that does not move on the receptor surface.



(a)



(b)



(c)

Figure 5. Scan paths to static patterns (a-c).

To demonstrate the attractor mechanism of saccade targeting, the pattern of figure 6 was modified to increase the contrast of the eyes and mouth by a factor of three relative to the remainder of the pattern (which was twice as bright as the background). The resulting

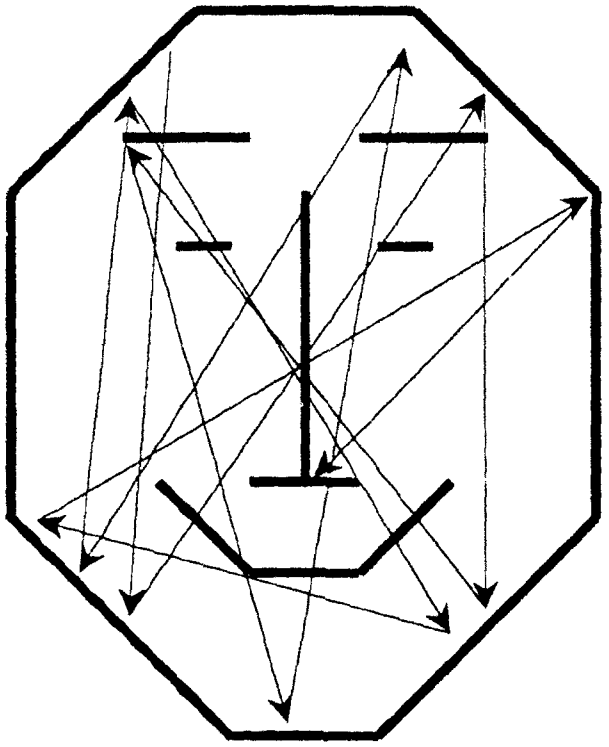


Figure 6. Scan path to a static image that was twice the size of those represented *in figures 4 and* 5.

scan path to the static pattern (figure 7) clearly shows a limit cycle involving the three targets. Return saccades to the mouth, which represents a more attractive target due to its size, are prevented by the mutual inhibition between homologous locations in the two bilateral SC until after another target has been selected.

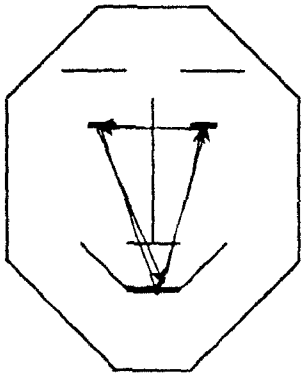


Figure 7. Scan path to static pattern when selected features are enhanced. The brightness of eyes and mouth are three times the brightness of the remainder of the pattern.

REFERENCES

Ali, M.A., and M.A. Klyne. (1985). *Vision in Vertebrates*. New York: Plenum Press.

Blackburn, M.R. (1993). "A simple computational model of center-surround receptive fields in the retina." TD 2454, NCCOSC RDT&E Division, San Diego, CA.

Blackburn, M.R., H.G. Nguyen, and P.K. Kaomea. (1987). "Machine visual motion detection modeled on vertebrate retina." *SPIE Proceedings*, vol. 980, pp. 90-98.

Breitmeyer, B.G. (1983). "Sensory masking, persistence, and enhancement in visual exploration and reading." In K. Rayner (ed.) *Eye movements in reading. Perceptual and Language Processes*. New York: Academic Press, pp. 3-30.

Cleland, B.G., M.W. Dubin, and W.R. Levick. (1971). 'Sustained and transient neurons in the cat's retina and lateral geniculate nucleus.' *Journal of Physiology*, vol. 217, pp. 473-496.

Cynader, M., and N. Berman. (1972). "Receptive field organization of monkey superior colliculus." *Journal of Physiology*, vol. 35, pp. 187-219.

Ewert, J.-P. (1978). "Neuronal basis of configurational prey selection in the common toad." In D.J. Ingle, M.A. Goodale, and J.W.R. Mansfield (eds.). *Analysis of Visual Behavior*. Cambridge, MA: MIT Press, pp. 7-45.

Fite, K.V., and F. Scalia. (1976). "Central visual pathways in the frog." In K.V. Fite (ed.). *The Amphibian Visual System: A Multidisciplinary Approach*. New York: Academic Press, pp. 87-118.

Lee, C., W.H. Rohrer, and D.L. Sparks. (1988). "Population coding of saccadic eye movements by neurons in the superior colliculus." *Nature*, vol. 332, pp. 357-360.

Perry, V.H., R. Oehler, and A. Cowey. (1984). "Retinal ganglion cells that project to the superior colliculus and pretectum in the macaque monkey." *Neuroscience*, vol. 12, pp. 1125-1137.

Rashbass, C. (1961). "The relationship between saccadic and smooth tracking eye movements." *Journal of Physiology*, vol. 159, pp. 326-338.

Sterling, P. (1983). "Microcircuitry of the cat retina." *Annual Review of Neuroscience*, vol. 6, pp. 149-185.

Straschill, M., and P. Rieger. (1973). "Eye movements evoked by focal stimulation of the cat's superior colliculus." *Brain Research*, vol. 59, pp. 221-227.

Wehner, R. (1981). "Spatial vision in arthropods." In H. Autrum (ed.) *Handbook of Sensory Physiology, VII/6C, Vision in Invertebrates.* Berlin: Springer-Verlag, pp. 287-616.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	February 1993	Final

4. TITLE AND SUBTITLE

MACHINE VISUAL TARGETING MODELED ON BIOLOGICAL REFLEXES

5. FUNDING NUMBERS

PE: 0602936N
WU: DN303020

6. AUTHOR(S)

M. R. Blackburn

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Command, Control and Ocean Surveillance Center (NCCOSC)
RDT&E Division
San Diego, CA 92152–5000

8. PERFORMING ORGANIZATION REPORT NUMBER

TD 2455

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Chief of Naval Research
Arlington, VA 22217

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The report presents a simple algorithm for machine visual target detection and localization based on aspects of the known functional neuroanatomy of biological systems. The system can generate, from the relative motion of a target image on the sensor surface, a series of motor commands that reorient the high-resolution portion of the sensor surface to regions of the target that potentially contain the most information. The reorientations are analogous to saccadic eye movements of biological visual systems.

14. SUBJECT TERMS

motion analysis
superior collicus
scan paths
scale invariance
saccadic eye movements

15. NUMBER OF PAGES

21

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAME AS REPORT

21a. NAME OF RESPONSIBLE INDIVIDUAL	21b. TELEPHONE *(Include Area Code)*	21c. OFFICE SYMBOL
M. R. Blackburn	(619) 553-1904	Code 943

INITIAL DISTRIBUTION

Code 0012	Patent Counsel	(1)
Code 0244	V. Ware	(1)
Code 943	M. R. Blackburn	(100)
Code 961	Archive/Stock	(6)
Code 964B	Library	(2)

Defense Technical Information Center
Alexandria, VA 22304-6145 (4)

NCCOSC Washington Liaison Office
Washington, DC 20363-5100

Center for Naval Analyses
Alexandria, VA 22302-0268

Navy Acquisition, Research & Development
Information Center (NARDIC)
Washington, DC 20360-5000

GIDEP Operations Center
Corona, CA 91718-8000

NCCOSC Division Detachment
Warminster, PA 18974-5000